Kashaf

Privacy Policy Checker

By : Mohammed Shammout

Alaa Bishtawi

Supervisors :Dr. Mona Demaidi

Dr. Samer Arandi

The Problem

- Privacy policies are long and full of technical terms; this makes reading them hard and tedious for ordinary people who want to access the service they want to use, resulting in many people ignoring those terms and accepting them in haste.
- Some companies and service providers use this in their advantage by requesting for more permissions and collect more data about the users.
- This puts the privacy and security of the users at risk.

Objective

- Our objective is to build a website to automatically evaluate English privacy policies and provide a general score for the whole document and a sentence-by-sentence score to highlight important notes for the user.
- We also aim to collect data to extend our data set by providing the users with the ability to rate sentences in the privacy policy.

Dataset

- To train our model we used the data from “tosdr.org”.
- The dataset contains ~15K sample. Each sample contains a sentence text, a score and a result.
- The dataset contained a lot of noise and errors that we had to deal with before training.

Dataset - Noise

- Non English texts:

  “Изменение настроек может…иональность сайта.\n”

- Noise in texts:

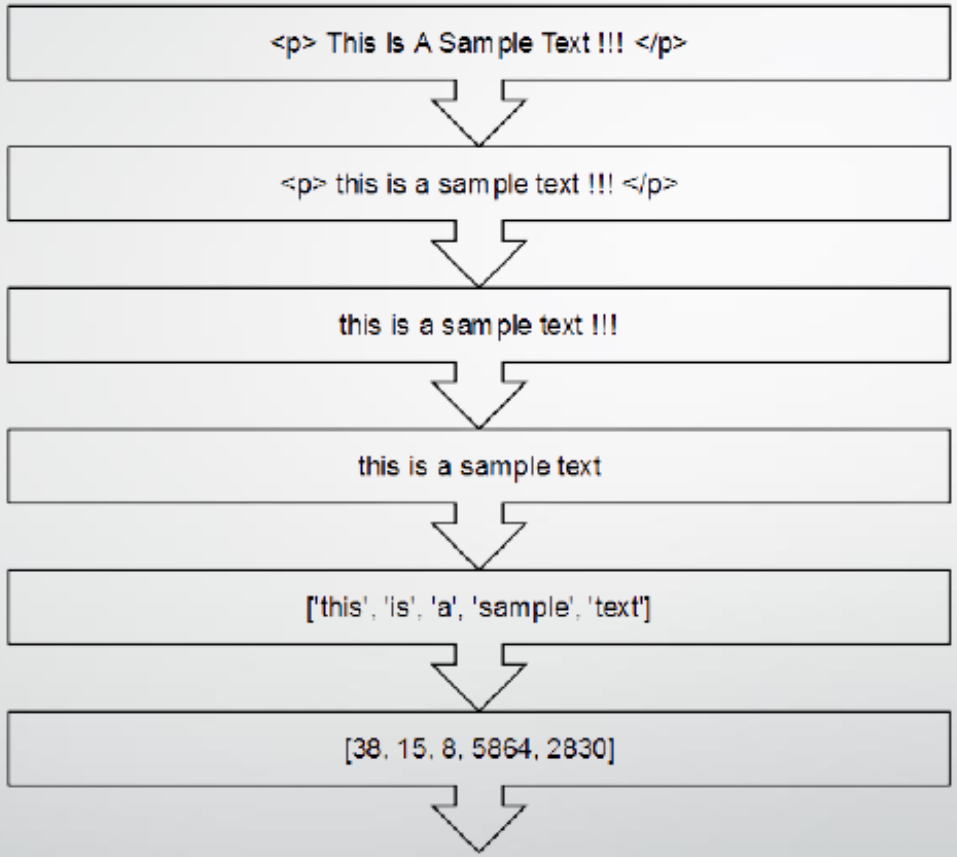
  “<div class=\"browser-compatibility-alert\">.<i class=\"fa fa-exclamation-triangle fa-2x pull-left\">.<p class=\"browser-error\">”

- Non suitable ratings:

  Many good points were tagged as bad and vice versa.

text Preprocessing

1. convert the text to lower case.
2. removing all HTML tags.
3. strip all punctuation from the text
4. tokenize the sentence to words.
5. indexing the words with the help of the embedding matrix.

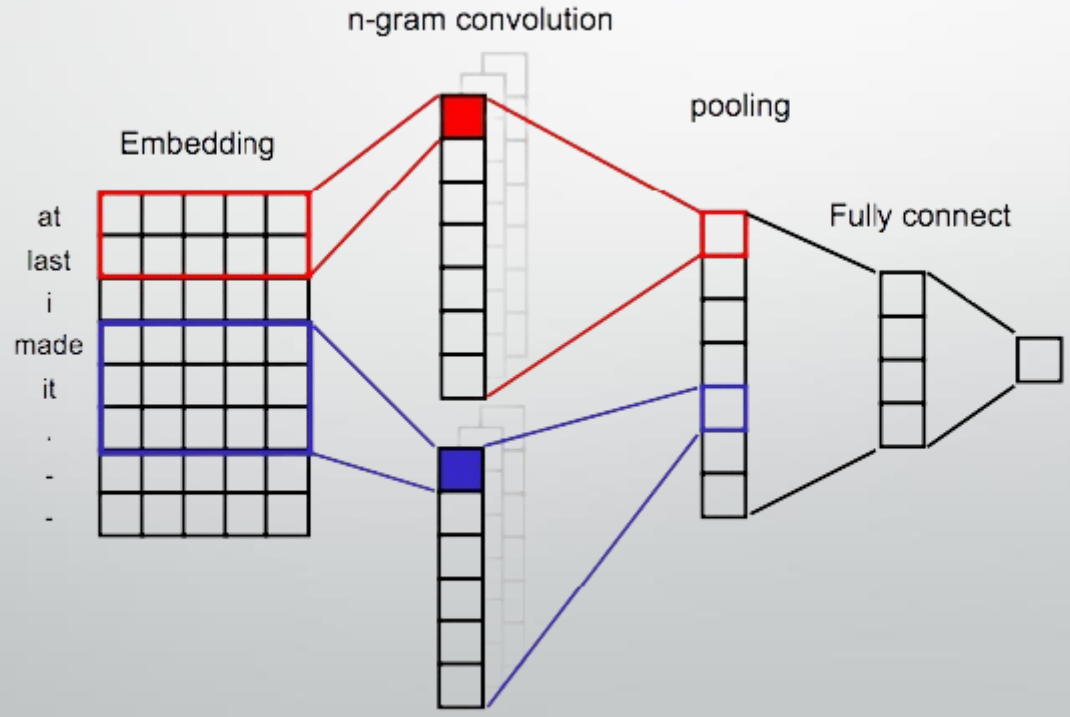

 This Is A Sample Text !!!
 this is a sample text !!!
this is a sample text !!!
this is a sample text
['this', 'is', 'a', 'sample', 'text']
[38, 15, 8, 5864, 2830]

Output

- the data set contains two attributes for each point:
  - class witch categorize the point to one of "good","natural","bad","blocker“
  - score which is an integer value in rage [0,100] that descries how important the point in general.
- We combined the class and score into one value ‘result’ witch is in range between -1 (very bad) and 1 (very good).

Model Architecture.

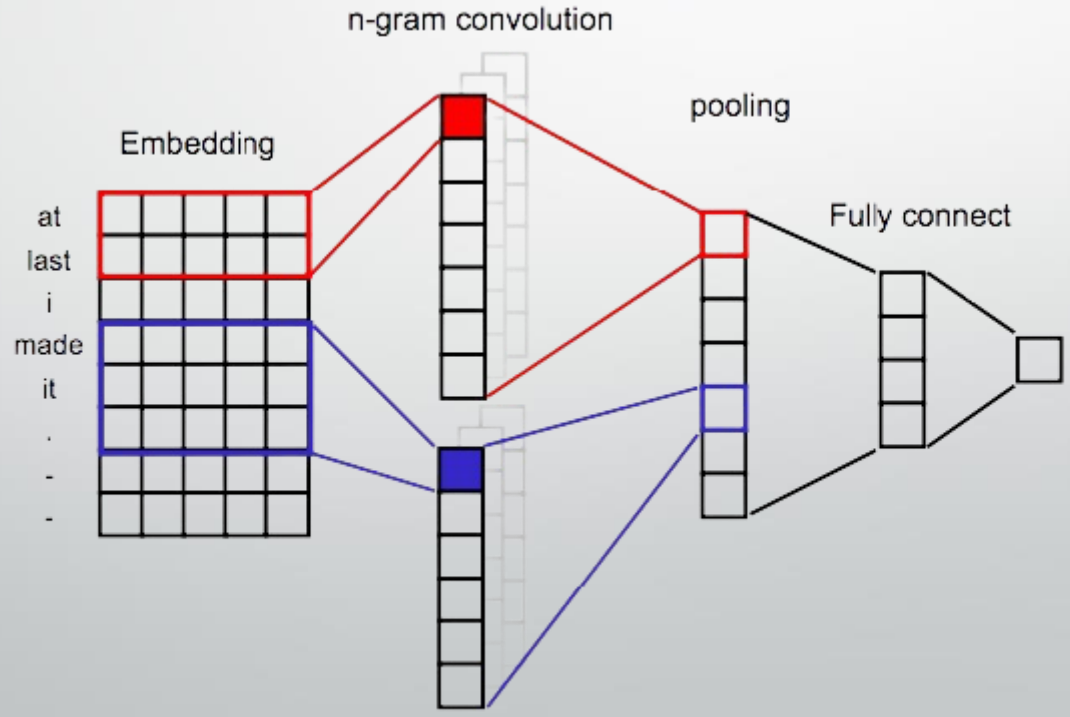
- We used word embedding with convolutional neural networks.



Layers:

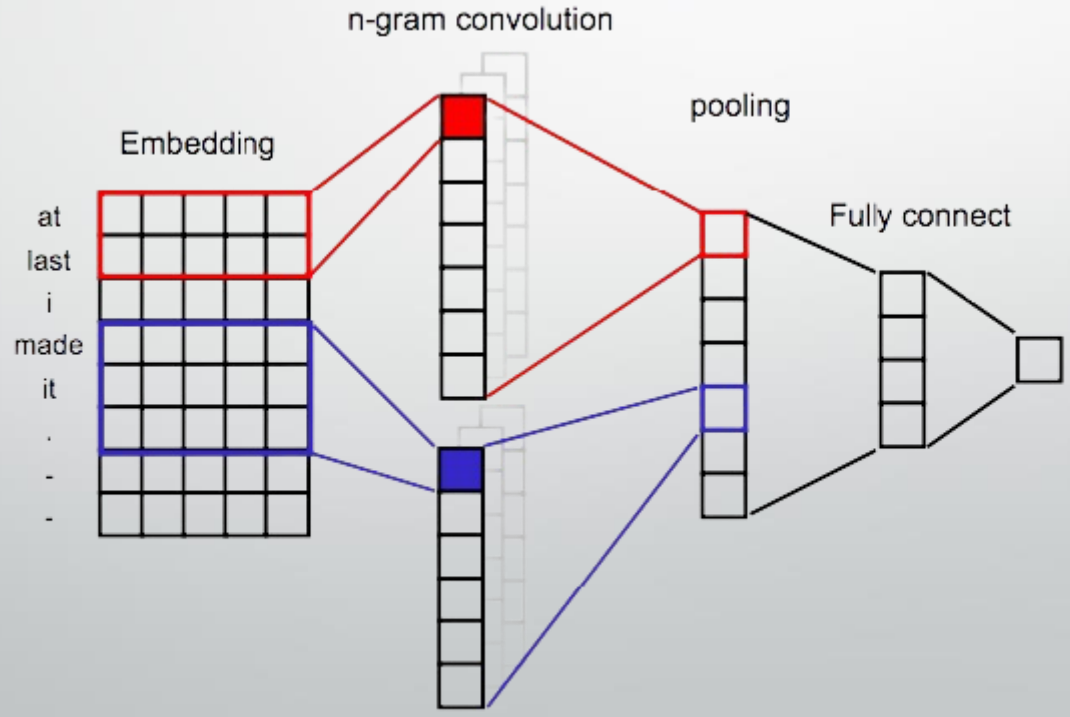
Word Embedding Layer: we used it to convert each word in the text to a vector of numerical values.

We used GloVe pre-trained embedding.



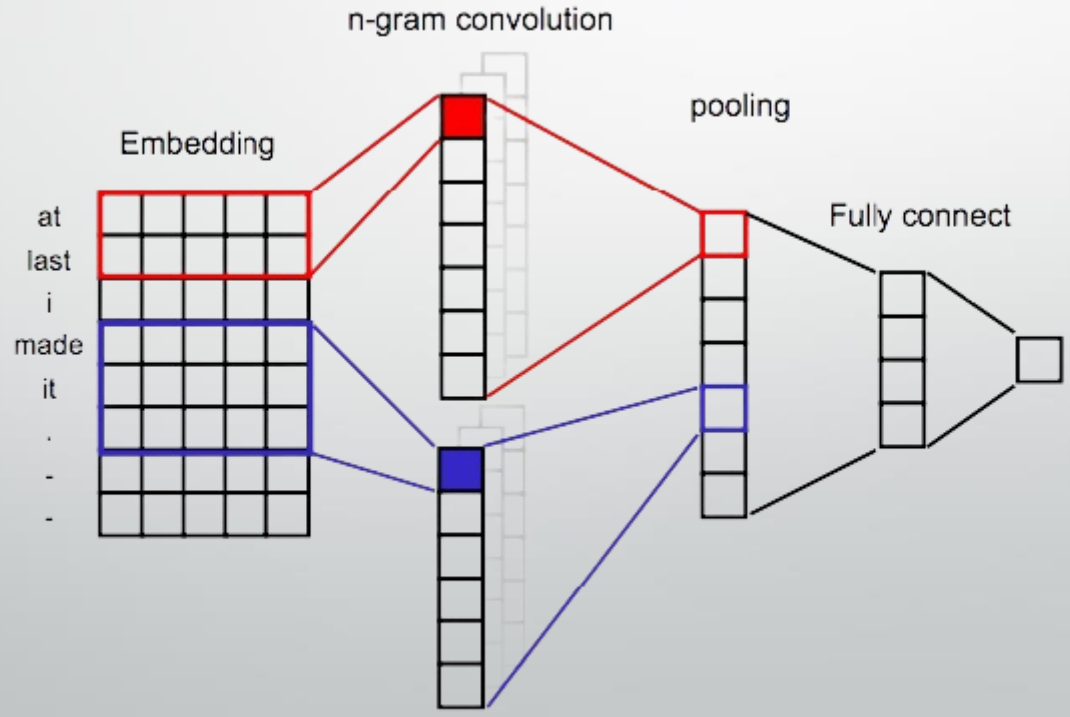
Layers:

1-Dimensional Convolutional Layer to extract fixed-size n-gram features



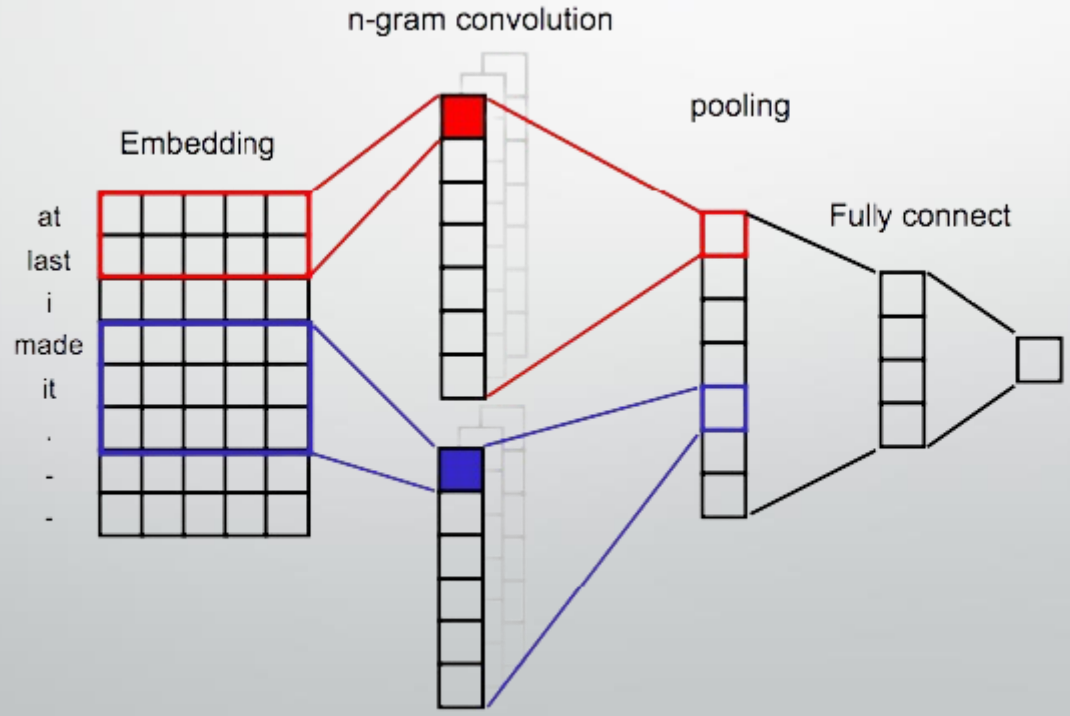
Layers:

Global average pooling: this layer calculates the average of each filter.



Layers:

Fully connected multi layer neural network that ends with the output layer.

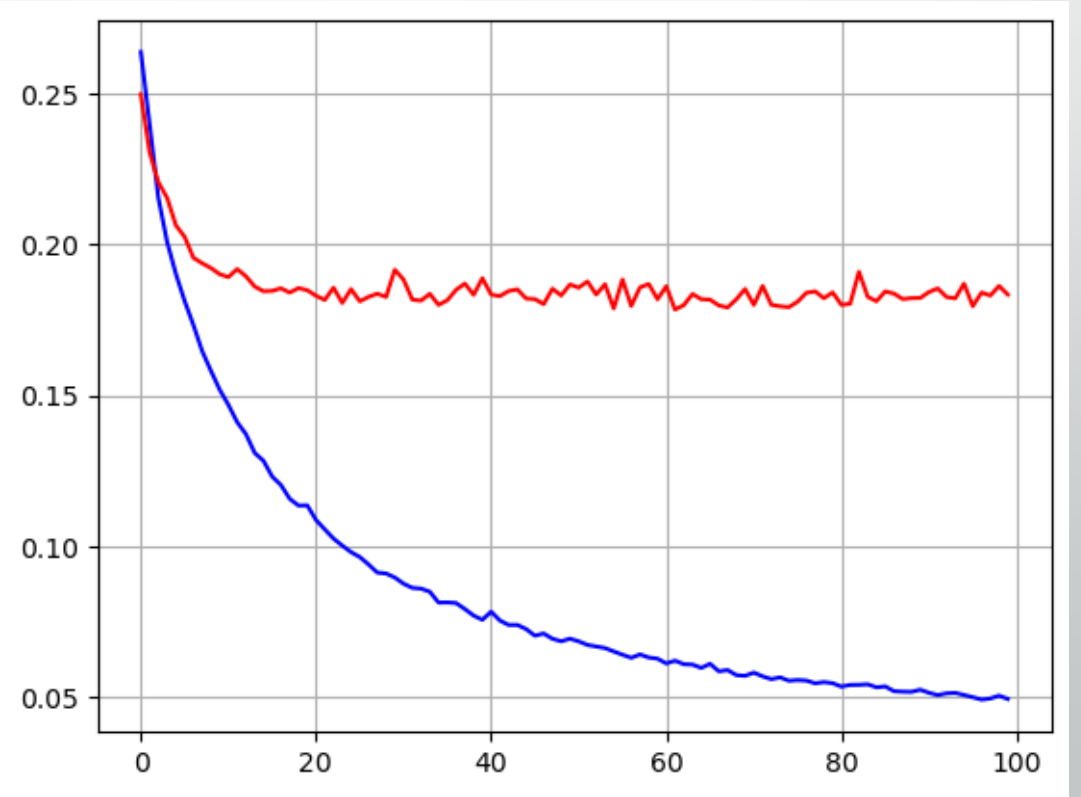


Results

- Our model suffered from over-fitting, this is because of the small dataset size for such a problem.

Mean square error

	<0.25	<0.20	<0.10
Training	96%	94%	89%
Validation	76%	70%	58%



Future Work

- Collect more data from the users. to solve the problem of over-fitting we want to give the users the ability to rate sentences the think they should be rated differently. The users can also provide us with a reason for the rating of the sentences to expand our model.
- Improve the model's architecture and try and experiment with other techniques.